Multicenter Study of Neurodevelopment in 3-Year-Old Children With and Without Single-Suture Craniosynostosis

Jacqueline R. Starr, PhD; Brent R. Collett, PhD; Rebecca Gaither, PhD; Kathleen A. Kapp-Simon, PhD; Mary Michaeleen Cradock, PhD; Michael L. Cunningham, MD, PhD; Matthew L. Speltz, PhD

Objective: To evaluate the hypothesis that 3-year-old children with single-suture craniosynostosis would receive lower neurodevelopmental scores than a comparable group of children born with patent sutures.

Design: Longitudinal comparison study.

Setting: Five tertiary care craniofacial centers.

Participants: Patients with craniosynostosis (cases) and a comparison group of children without craniosynostosis (controls). Patients diagnosed with single-suture craniosynostosis from 2002 to 2006 were eligible as cases. Controls were frequency-matched to cases on age, sex, race, socioeconomic status, and study site.

Main Exposure: Craniosynostosis.

Main Outcome Measures: We administered the Bayley Scales of Infant Development, Second Edition, mental and motor development indices and the Preschool Language Scales, Third Edition, receptive and expressive communication scales. Children were evaluated at baseline (before surgery in cases and at a similar age in controls) and at 18 and 36 months of age. We compared the groups’ performances at 36 months by fitting adjusted linear and logistic regression models. We also estimated adjusted associations between age at surgery and neurodevelopmental scores.

Results: Adjusted mean case deficits ranged from 3 to 6 points (P ≤ .008 for all comparisons). Compared with controls, the odds of cases being delayed ranged from 1.5 to 2.0, depending on the neurodevelopmental scale (P values ranged from .03 to .09). Cases’ ages at craniosynostosis repair were not strongly related to neurodevelopmental performance.

Conclusions: In this large, carefully controlled, multicenter study, we observed consistently lower mean neurodevelopmental scores in children with single-suture craniosynostosis compared with controls. These results provide further support for neurodevelopmental screening in young children with single-suture craniosynostosis.

SINGLE-SUTURE CRANIOSYNOSTOSIS (SSC), the premature fusion of one of the cranial sutures, occurs in approximately 1 in 1700 to 2500 live births.1,2 A longstanding question regarding the sequelae of SSC has been whether the diagnosis increases children’s risk of neurodevelopmental problems,3 in part because developmental delay is a feature of some syndromes involving craniosynostosis.4-6 Furthermore, premature suture fusion constrains the skull during a time of rapid brain growth, alters brain morphology,7,8 and may cause increased intracranial pressure. It is also possible that craniosynostosis is due to a primary brain defect that influences the timing of suture fusion.7,10 Regardless of whether potential developmental delays are caused by SSC or concomitantly result from a risk factor common to both conditions, the confirmation of this association will inform neurodevelopmental screening and intervention efforts.

Accumulating evidence suggests that most children with isolated SSC score in the normal range on neurodevelopmental examinations. Collectively, however, their average scores are lower than those of children with patent sutures, suggesting an elevated risk of developmental delays.6,10-12 Early investigations of this relationship often used weak study designs with unclear participant eligibility criteria, smaller sample sizes, or nonstandardized assessments of neurodevelopment, and most lacked comparison groups of children born with patent sutures.3,10 We previously published interim results of a

Author Affiliations are listed at the end of this article.
A longitudinal study designed to overcome these limitations. Herein, we present comparisons of cognitive, motor, and language development in 36-month-old children with and without SSC who were recruited from 5 craniofacial centers across the United States.

STUDY DESIGN

We conducted a multicenter, longitudinal study of children with SSC (cases) and children without craniosynostosis (controls); the controls were frequency-matched to cases on the basis of sex, age at baseline, race/ethnicity, and socioeconomic status (SES). We sought to enroll all eligible infants between January 2002 and September (cases) or December (controls) 2006 from Seattle Children’s Hospital (Washington); the Children’s Memorial Hospital through Northwestern University in Chicago, Illinois; Children’s Health Care of Atlanta (Georgia); and St Louis Children’s Hospital (Missouri). We also approached eligible children diagnosed at Children’s Hospital of Philadelphia (Pennsylvania) starting in January 2006 (who were included in the Chicago site’s numbers because they were assessed by the Chicago team).

CASES

Infants with SSC were referred to the project at the time of diagnosis by the treating surgeon or pediatrician. Infants were eligible if they (1) had SSC (sagittal, metopic, unilateral coronal, or unilateral lambdoid synostosis), confirmed by computed tomography scans; (2) had not yet had reconstructive surgery; and (3) were 30 months of age or younger at recruitment. Exclusion criteria for cases included (1) prematurity (<34 weeks’ gestation); (2) major medical or neurological conditions (eg, cardiac defects, seizure disorders, or significant health conditions requiring surgical correction); (3) the presence of 3 or more extracranial minor malformations; or (4) the presence of other major malformations.

We identified 333 cases, of whom 322 were deemed eligible. The parents of 52 cases declined either passively, by not responding to attempts to contact them (n=23), or actively, by expressing lack of interest or time (n=29). We enrolled 270 cases (84% of those eligible), 4 of whom were later found to be ineligible. Among the 266 cases seen at baseline, 210 had a 36-month study visit, 209 with valid outcome data.

CONTROLS

Infants were eligible as controls if they had no known craniofacial anomaly and met none of the exclusion criteria for cases. Control group participants were recruited through pediatric practices, birthing centers, and announcements in publications of interest to parents of newborns. Controls were frequency-matched to cases in relation to factors potentially related to neurodevelopmental performance and craniosynostosis risk (ie, potential confounders): (1) age at enrollment (within ±3 weeks); (2) sex; (3) family SES within the same Hollingshead category; and (4) race/ethnicity. We approached the interested parents of 347 eligible controls who matched to enrolled cases. The 262 enrolled participants (3 were later declared ineligible) represented 76% of all those who were invited to participate as controls. Half those who declined (n=45) did so passively and half actively (n=40). Among the 259 controls seen at baseline, 224 had a 36-month study visit, 222 with valid outcome data.

There were 2 related pairs of cases and 2 related pairs of controls; eliminating one of each pair did not appreciably alter the results. We offered $100 compensation for each study visit.

MEASURES

We used the Bayley Scales of Infant Development, Second Edition (BSID-II), to measure infants’ cognitive and psychomotor status. The BSID-II is a frequently used, standardized, validated, norm-referenced objective test of young children’s developmental status from 16 days to 42 months, 15 days of age. It yields separate indices of mental and psychomotor development: the Mental Development Index (MDI) and Psychomotor Development Index (PDI). We used the Preschool Language Scale. Third Edition (PLS-3), to assess expressive and receptive language skills. The PLS is a norm-referenced, validated, individually administered objective test of language in young children. It yields 2 scale scores, receptive (PLS-AC) and expressive (PLS-EC) communication scales, and a total language score. Norms are provided for infants and preschoolers from 2 weeks to 83 months.

For both the BSID-II and PLS-3, standard scores are derived for all scales with a normative average of 100 and standard deviation of 15. All BSID-II and PLS-3 assessments were performed by trained psychometrists and videotaped for reliability purposes. We have previously described assessments of reliability, for which agreement on individual items was 98% or more for the MDI and both PLS-3 scales and 93% for the PDI.

PROCEDURES

We obtained informed consent following the institutional review board–approved protocols of each participating institution. There were 3 study visits: before surgery (for cases, and at a similar age for controls) and at approximately 18 and 36 months of age. At each visit, psychometrists administered the BSID-II and the PLS-3, and we interviewed mothers regarding children’s medical history. At the first visit, mothers also independently completed an intelligence test, the Wonderlic Personnel Test. This report pertains to neurodevelopmental assessments made at the third study visit.

DATA ANALYSES

Unless otherwise specified, all regression analyses involved robust estimation of standard errors and 95% CIs and were adjusted for sex; family SES (Hollingshead composite score, continuous); age at assessment (in months); race/ethnicity (white or other); maternal IQ; and recruitment site. To estimate the differences in mean test scores between cases and controls at the third visit, we performed linear regression analyses with each neurodevelopmental outcome in separate adjusted analyses.

To evaluate whether cases’ test scores varied by diagnosis group (ie, site of the fused suture), we fit a linear regression model, including indicators of diagnostic group, and estimated an overall Wald P value for the group. We also compared the proportion of cases and controls testing in the “delayed” range (ie, standardized scores <85) by fitting logistic regression models.

To explore potential biases in the data, we repeated the primary linear regression analyses after excluding, in separate analyses, 25 cases in whom mutations were detected through the sequencing of exons and intron-exon boundaries in craniosynostosis syndrome–causing genes; 53 cases and 31 controls who received intervention services, such as speech or development-
At 36 months, the cases’ and controls’ distributions of demographic characteristics differed little from those in the original sample, again, approximately one-third of participants were girls, and more than 70% were of white, non-Hispanic race/ethnicity (Table 1). Attrition was not greatly related to study center but was predicted by SES, with the proportion in the lowest SES categories decreasing from 16% to 10%. Half the subjects were 36 months, and half were older, ranging up to 43 months. Cases were lost to follow-up more often than controls (21% vs 14%), although, among cases, attrition did not differ by the affected suture.

Controls’ median standardized neurodevelopmental scores were approximately 100 or greater on all scales (MDI, PLS-AC, and PLS-EC) except the PDI, for which the median score was 95. Controls’ mean scores were generally higher for participants who were female, white, and of higher SES (eTable). Cases’ mean scores were lower than those of controls’ on all 4 scales (Table 2). In linear regression analyses, adjusted mean case deficits ranged from 3 to 6 points (P < .008 for all comparisons).

Among cases, children with sagittal synostosis had the highest neurodevelopmental scores, on average. Compared with them, adjusted mean scores among children with metopic synostosis ranged from 0.9 to 4.3 points worse, whereas the range for children with lamboid synostosis was 8 to 14 points worse, depending on the neurodevelopmental scale (Table 3).

For 3 of the 4 outcome scales, 10% to 13% of controls scored in the delayed range, as did 22% to 25% of cases (Table 4). The greatest proportion of children were delayed on the PDI (~20% of controls and ~28% of cases). Compared with controls, the odds of cases’ being delayed ranged from 1.3 to 2.0 depending on the neurodevelopmental scale (P values ranged from .03 to .09).

Excluding subgroups of participants for various reasons in separate analyses led to a lessening of estimated case-control differences, generally from 0.5 to 1 point in mean test scores, and increases in P values. For example, after excluding cases with genetic mutations, the remaining cases now lagged controls, on average, by only 3.3 points on the MDI (vs 4.2 points when all cases were included). Applying inverse probability weighting to linear regression analyses yielded estimated case deficits that were very close to and generally larger than those in the primary analyses (data not shown).

Three children had not had surgery by the third visit. Of the remaining participants, children with sagittal synostosis tended to undergo earlier synostosis repair, at a median of 5 months vs 9.5 months for children with nonsagittal synostosis. The ages at surgery ranged from 2 months to 3 years. Children in families of higher SES also tended to undergo synostosis repair at younger ages. On average, for the MDI, PLS-AC, and PLS-EC, children scored 0.3 points worse in association with each 1-month increase in age at surgery (all P > .13; Table 5). These differences attenuated slightly after adjusting for presurgery scores on the corresponding BSID-II or PLS-3 scale and when we incorporated inverse probability weights into the analysis (all P > .15). The PDI scores were unrelated to age at surgery.

### RESULTS

At 36 months, the cases’ and controls’ distributions of demographic characteristics differed little from those in the original sample, again, approximately one-third of participants were girls, and more than 70% were of white, non-Hispanic race/ethnicity (Table 1). Attrition was not

### COMMENT

Consistent with the differences that we observed before cases’ surgical procedures and around the age of 18

### Table 1. Demographic Characteristics of 3-Year-Old Children With (Cases) and Without (Controls) Single-Suture Craniosynostosis Evaluated at the Third of 3 Study Visits

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>209 (100.0)</td>
<td>222 (100.0)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>136 (65.1)</td>
<td>139 (62.6)</td>
</tr>
<tr>
<td>Female</td>
<td>73 (34.9)</td>
<td>83 (37.4)</td>
</tr>
<tr>
<td>Age group, mo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;37</td>
<td>104 (49.8)</td>
<td>110 (49.5)</td>
</tr>
<tr>
<td>≥37</td>
<td>105 (50.2)</td>
<td>112 (50.5)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonwhite</td>
<td>49 (23.4)</td>
<td>64 (28.8)</td>
</tr>
<tr>
<td>White</td>
<td>160 (76.6)</td>
<td>158 (71.2)</td>
</tr>
<tr>
<td>SES category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>50 (23.9)</td>
<td>61 (27.5)</td>
</tr>
<tr>
<td>2</td>
<td>76 (36.4)</td>
<td>115 (51.8)</td>
</tr>
<tr>
<td>3</td>
<td>50 (23.9)</td>
<td>28 (12.6)</td>
</tr>
<tr>
<td>4</td>
<td>20 (9.6)</td>
<td>10 (4.5)</td>
</tr>
<tr>
<td>5</td>
<td>10 (4.8)</td>
<td>4 (1.8)</td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seattle</td>
<td>72 (34.4)</td>
<td>80 (36.0)</td>
</tr>
<tr>
<td>Chicago</td>
<td>81 (38.8)</td>
<td>87 (39.2)</td>
</tr>
<tr>
<td>St Louis</td>
<td>26 (12.4)</td>
<td>20 (9.0)</td>
</tr>
<tr>
<td>Atlanta</td>
<td>30 (14.4)</td>
<td>35 (15.8)</td>
</tr>
</tbody>
</table>

Abbreviation: SES, socioeconomic status.
months, 36-month-old children treated for SSC tended to score lower on neurodevelopmental assessments than a comparable group of children without SSC. Although the observed mean differences of 3 to 6 points, or approximately one-quarter to one-half a sample standard deviation, may appear small, cases had 1.5 to twice the...
odds of scoring in the delayed range on the 4 cognitive, motor, and speech scales.

These differences also persisted in the multiple sub-analyses that we conducted to explore potential biases. They do not appear attributable to undiagnosed syndromes, although additional cases may harbor as-yet unidentified genetic or epigenetic differences that contributed to both SSC and neurodevelopment. Excluding participants on the basis of other factors, including medical histories of seizures, head injuries, or surgeries, also did not alter the interpretation of the primary results.

The validity of the study findings rests on the assumption that controls are indeed comparable to cases. Through frequency-matching, we enrolled case and control groups that were similar in their distributions of demographic characteristics, including age at testing, sex, familial SES, race/ethnicity, and study site. Control participants may have never been definitive evidence to indicate that this correlation between duration of anesthesia (or surgery) and subsequent neurodevelopmental test scores among the Seattle site patients in this study (manuscript under review). Due to the inextricability of surgery duration with patient characteristics, such as bone thickness and difficulty of the craniostenosis repair, these results must be considered preliminary. Nevertheless, they offer one possible causal pathway for the observed neurodevelopmental differences between patients with SSC and controls, in addition to others previously hypothesized, such as genetic factors or intracranial pressure.

Early surgery helps to normalize children’s skull shape and appearance and also minimizes the chance of complications due to increased intracranial pressure. It has been postulated that early surgery will also decrease the chance of neurodevelopmental deficits potentially caused either by increased intracranial pressure or by constraint of brain growth and consequent structural differences. There has never been definitive evidence to indicate that this concern should influence the timing of surgery. Designing a US study to address this question conclusively is challenging, if not impossible, because so few parents of infants diagnosed with SSC forgo surgery.

We and others have, instead, examined the timing of surgery in relation to neurodevelopment, with later surgical procedures serving as a rough surrogate for an unoperated group. We observed only a weak, if any, relationship at 36 months and at an earlier postsurgery assessment. Although the data do not refute an inverse relationship between age at surgery and neurodevelopment, they also do not provide strong evidence that such a relationship exists.

There has also been ongoing interest in determining whether SSC-related neurodevelopmental problems differ depending on the affected suture. With assessments at earlier ages, we observed inconsistent differences by suture group that were difficult to interpret. At the latest, 36-month assessments, suture group differences were more consistent across the various neurodevelopmental scales, with patients with sagittal synostosis tending to

<table>
<thead>
<tr>
<th>Neurobehavioral Test</th>
<th>Children, No.</th>
<th>Mean Differencea (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayley Scales of Infant Development</td>
<td>Mental Development Index</td>
<td>203</td>
<td>−0.3 (−0.7 to 0.1)</td>
</tr>
<tr>
<td>Bayley Scales of Infant Development</td>
<td>Psychomotor Development Index</td>
<td>202</td>
<td>0.0 (−0.3 to 0.4)</td>
</tr>
<tr>
<td>Preschool Language Scale</td>
<td>Receptive communication</td>
<td>200</td>
<td>−0.3 (−0.7 to 0.2)</td>
</tr>
<tr>
<td>Preschool Language Scale</td>
<td>Expressive communication</td>
<td>200</td>
<td>−0.3 (−0.6 to 0.1)</td>
</tr>
<tr>
<td>Preschool Language Scale</td>
<td>Total</td>
<td>200</td>
<td>−0.3 (−0.6 to 0.1)</td>
</tr>
</tbody>
</table>

*Adjusted for suture group (sagittal, metopic, right or left unicoronal combined, or lambdoid); familial socioeconomic status (Hollingshead 4-factor classification, included as a continuous score); race/ethnicity (white vs other); maternal IQ (continuous); and recruitment site.

In addition, we conducted inverse probability-weighted analyses that accounted for some possible biases induced by differential loss to follow-up. These analyses produced stronger estimated differences between cases and controls. As in most longitudinal studies, it is possible that there are remaining unmeasured factors that could be related to attrition, case status, and neurodevelopmental test scores. Nevertheless, it is unlikely that there exists such an unknown factor with strong enough relationships to have rendered the observed case-control differences wholly artifactual, particularly given that only 18% of subjects missed the 36-month visit.

In the United States, children typically undergo cranial surgery to release the fused suture in the first year of life. Cases thus differ from the vast majority (if not all) of the controls in having undergone a major surgical procedure, with concomitant anesthesia exposure. There is compelling evidence from animal studies that exposure to anesthetic agents can damage the developing brain. Some studies suggest that repeated or prolonged exposure to anesthesia in early life may similarly affect human development. Indeed, we observed an inverse relation between duration of anesthesia (or surgery) and subsequent neurodevelopmental test scores among the Seattle site patients in this study (manuscript under review). Due to the inextricability of surgery duration with patient characteristics, such as bone thickness and difficulty of the craniostenosis repair, these results must be considered preliminary. Nevertheless, they offer one possible causal pathway for the observed neurodevelopmental differences between patients with SSC and controls, in addition to others previously hypothesized, such as genetic factors or intracranial pressure.

Early surgery helps to normalize children’s skull shape and appearance and also minimizes the chance of complications due to increased intracranial pressure. It has been postulated that early surgery will also decrease the chance of neurodevelopmental deficits potentially caused either by increased intracranial pressure or by constraint of brain growth and consequent structural differences. There has never been definitive evidence to indicate that this concern should influence the timing of surgery. Designing a US study to address this question conclusively is challenging, if not impossible, because so few parents of infants diagnosed with SSC forgo surgery.

We and others have, instead, examined the timing of surgery in relation to neurodevelopment, with later surgical procedures serving as a rough surrogate for an unoperated group. We observed only a weak, if any, relationship at 36 months and at an earlier postsurgery assessment. Although the data do not refute an inverse relationship between age at surgery and neurodevelopment, they also do not provide strong evidence that such a relationship exists.

There has also been ongoing interest in determining whether SSC-related neurodevelopmental problems differ depending on the affected suture. With assessments at earlier ages, we observed inconsistent differences by suture group that were difficult to interpret. At the latest, 36-month assessments, suture group differences were more consistent across the various neurodevelopmental scales, with patients with sagittal synostosis tending to...
score the highest, patients with metopic synostosis performing worse than but similarly to patients with sagittal synostosis, and the small group of patients with lambdoid synostosis consistently scoring the lowest. Because of the smaller sample sizes in these subdivided analyses, these analyses should be considered exploratory.

The clinical significance of the study findings is presently unclear. The magnitude of group differences and the fact that most patients scored in the average to low average range of test norms suggest that children with SSC may be at risk for “high prevalence/low severity problems” at school age, including learning disabilities and academic underachievement. Confirmation that, indeed, these school-age problems do occur more frequently in children with SSC would further justify early developmental screening because such problems may be particularly amenable to early intervention. However, the early detection of subtle neurodevelopmental risk is notoriously difficult, and previous studies have shown the predictive validity of the BSID-II to be variable, depending on children’s age and other characteristics. Prediction generally improves when children are 24 months or older and when assessments include language functions such as those measured by the PLS-3. Nevertheless, the predictive function of the BSID-II and PLS for children with SSC will remain unclear until we complete our assessments of this cohort at age 7 years, assessments that are currently under way.

Three-year-old children with SSC scored lower on neurodevelopmental assessments than did a comparable group of children without SSC, showing persistence of an observed trend for mildly delayed functioning since the first year of life, before cranial vault surgery. Although we do not yet know what explains the association between SSC and delayed functioning, these findings suggest that the neurodevelopment of young children with SSC should be monitored closely.

Accepted for Publication: December 13, 2011.

Author Affiliations: Departments of Epidemiology (Dr Starr), Psychiatry and Behavioral Sciences (Drs Collett and Speltz), and Pediatrics (Dr Cunningham), University of Washington, and Children’s Craniofacial Center (Drs Cunningham and Speltz), and Psychiatry and Behavioral Medicine (Drs Collett and Speltz), Seattle Children’s Hospital, Seattle, Washington; The Forsyth Institute, Cambridge, Massachusetts (Dr Starr); Department of Surgery, Northwestern University, and Shriners Hospital for Children (Drs Gaither and Kapp-Simon); Chicago, Illinois; and Department of Psychology, St Louis Children’s Hospital, and Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri (Dr Cradock).

Correspondence: Matthew L. Speltz, PhD, Seattle Children’s Hospital, 4800 Sand Point Way, NE, Mailstop CL-08, Seattle, WA 98105 (mspeltz@uw.edu).

Author Contributions: Dr Speltz had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Starr, Kapp-Simon, Cradock, Cunningham, and Speltz. Acquisition of data: Starr, Collett, Gaither, Kapp-Simon, Cradock, Cunningham, and Speltz. Analysis and interpretation of data: Starr, Collett, Kapp-Simon, Cunningham, and Speltz. Drafting of the manuscript: Starr, Cunningham, and Speltz. Critical revision of the manuscript for important intellectual content: Starr, Collett, Gaither, Kapp-Simon, Cradock, Cunningham, and Speltz. Statistical analysis: Starr. Obtained funding: Kapp-Simon, Cunningham, and Speltz. Administrative, technical, and material support: Collett, Kapp-Simon, Cradock, and Speltz. Study supervision: Starr, Kapp-Simon, and Speltz.

Financial Disclosure: None reported.

Funding/Support: This work was supported by grant R01 DE 13813 from the National Institute of Dental and Craniofacial Research (awarded to Dr Speltz).

Online-Only Material: The eTable is available at http://www.archpediatrics.com.

Additional Contributions: We thank Lauren Buono, PhD, for her role as a coinvestigator in this project; Sharman Conner, MA, for project coordination; our data analysts, including Kristen Daniels, MLIS, and Kristen Gray, MS, and Diana Prise, BA, for assisting the interviewers and doing data entry and validation. We also thank the families who have graciously participated in this research.

REFERENCES

15. Leppig KA, Werler MM, Cann CI, Cook CA, Holmes LB. Predictive value of minor